International Journal of Botany
and Research (IJBR)
ISSN (P):2277-4815; ISSN (E):2319-4456
Vol. 8, Issue 4, Aug 2018, 1-10

ASSOCIATION OF SANDALWOOD WITH OTHER SPECIES IN PACHAIMALAI HILLS WITH RELIABILITY OF ESTIMATES AND REGRESSION

DELPHINSONIA. M[1], JOHN ROBINSON. P[2] & SEBASTIAN RAJASEKARAN. C[3]

[1,3]Department of Botany, Bishop Heber College, Trichy, Tamil Nadu, India

[2]Department of Mathematics, Bishop Heber College, Trichy, Tamil Nadu, India

ABSTRACT

This research work is a continuation of the study that was previously made on the association of sandalwood with neighboring species in Pachaimalai hills, Eastern Ghats, Tamilnadu, India. We add a regression model through reliability of estimates to the procedure in order to predict one data instance from contributing more than others. We also utilize statistical analysis for the data processing and the experimental results show a very high utility of modeling plant association problems. The reliability of estimates and regression analysis done in the work shows close associations and relationships between two plant species clustering around Sandalwood in almost all the area under study in Pachaimalai hills.

***KEYWORDS:** Regression Analysis; Reliability of Estimates; Plant Association & Pachaimalai Hills*

Received: May 11, 2018; **Accepted:** Jun 01, 2018; **Published:** Jun 30, 2018; **Paper Id.:** IJBRAUG20181

Original Article

INTRODUCTION

Pachaimalai, also known as the Pachais (Sonia et al., 2015), are hills which are part of Eastern Ghats in Tamil Nadu in Trichy district, located near 11°11′N 78°21′E / 11.18°N 78.35°E / 11.18; 78.35. They are much greener than some of the other hills in the vicinity. Extending over an area of about 5,200 square miles (13,500 square km), they form a discontinuous line of highlands with a general elevation from 1,770 to 4,620 feet (540 to 1,400 meters). The economy of the region is based on agriculture; rice, jowar (grain sorghum), sugarcane, gram (chickpeas), peanuts (groundnuts), and *bajra* (pearl millet) are subsistence crops. Coffee, cashews, and pepper are important plantation crops raised for export. Pachaimalai is known for its thick vegetation of Sandalwood and bamboo. Santalum album or Indian sandalwood is a small tropical tree, and is the most commonly known source of sandalwood. This species has historically been cultivated, processed and traded since ancient times. Certain cultures place great significance on its fragrant and medicinal qualities. The high value of the species has caused its past exploitation, to the point where the wild population is vulnerable to extinction, especially in Pachaimalai hills known for its rich vegetation in sandalwood. For a sandalwood tree, a fruit is produced after three years and viable seeds after five years and these seeds are distributed by birds. Plant association is defined as the grouping of plant species, or a plant community, that recurs across the landscape. Plant associations are used as indicators of environmental conditions such as temperature, moisture, light etc. It can be viewed as a collection of plant species within a designated geographical unit, which forms a relatively uniform patch, distinguishable from neighboring patches of different vegetation types (Willner, 2006; Yang et al., 2010; Schleicher et al., 2011). The components of

each plant community are influenced by soil type, topography, climate and human disturbance. In many cases, there are several soil types within a given phytocoenosis. Many researchers have applied data-mining and recent advancements in various environmental situations in different parts of the globe (Andreopoulou et al., 2011; Chaudry, 2013; Rojas-Mora et al., 2013; Selmaui-Folcher et al., 2011). By using the fuzzy correlation analysis, fuzzy partial correlation, fuzzy semi-partial correlation and fuzzy multiple correlation rules for fuzzy numbers are generated to see that two fuzzy sets not only frequently occur together in same records, but also are related to each other. Correlation measures and data mining algorithm with numerical results were discussed in detail in our earlier work (Sonia et al., 2015). In this paper, we attempt to find some useful relationships through reliability of estimates in theregression analysis for the same numerical results of the species clustering around sandalwood in Pachaimalai hills.

MATERIALS AND METHODS

Since the distribution of sandalwood in the Pachaimalai hills is not even or unpredictable, surveying the frequency of plant associations with sandalwood is a tedious job. Hence we identified twelve potential spots in the hills where the distribution of the same is considered notable and the surveyed frequency data is converted into fuzzy numbers due to the irregular distribution of the plant communities. Four different plant species are found to be closely associated with sandalwood and they seem to recur in all the surveyed spots. In this paper, we utilize bio-statistical rules and determine the association between sandalwood with the identified four other plant species. Fuzzy correlation rules, fuzzy partial correlation rules, fuzzy semi-partial correlation rules, fuzzy multiple correlation rules and fuzzy regression analysis (Han &Kamber, 2012; Robinson et al., 2012; Agrawal et al., 1993) are used for identifying the plant associations.

Case Study of Plant Association with Sandalwood in Pachaimalai Hills

Assume that $T = \{t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8, t_9, t_{10}, t_{11}, t_{12}\}$ is a random sample with 12 fuzzy records shown in Table 1, and $F = \{f_1, f_2, f_3, f_4, f_5\}$ is the set of observed fuzzy items. In this study, we observe five plant species occurring frequently almost in all transactions of a botanical survey conducted in Pachaimalai hills, the Eastern Ghats for analyzing plant associations for sandalwood. The four plant species found to be distributed around sandalwood in the hills are Acacia pennata, Celtisphilippensis, Zizyphusoenoplia, and Miliusaeriocarpa.

Table 1: A Random Sample with 12 Fuzzy Records

F \ T	F_1 Santalum album	F_2 Acacia Pennata	F_3 Celtisphilippensis	F_4 Zizyphusoenoplia	F_5 Miliusaeriocarpa
t_1	0.78	0.71	0.65	0.6	0.56
t_2	0.98	0.89	0.82	0.75	0.7
t_3	0.56	0.51	0.47	0.43	0.4
t_4	0.66	0.6	0.55	0.51	0.48
t_5	0.4	0.36	0.33	0.31	0.29
t_6	0.16	0.15	0.13	0.12	0.11
t_7	0.34	0.31	0.28	0.26	0.24
t_8	0.9	0.82	0.75	0.69	0.64
t_9	0.22	0.2	0.18	0.17	0.16
t_{10}	0.64	0.58	0.53	0.49	0.46
t_{11}	0.3	0.27	0.25	0.23	0.21
t_{12}	0.48	0.44	0.4	0.37	0.34

The data presented in table 1 is generated using the following fuzzy membership functions for the frequency of the five different plant species occurring together frequently.

Figure 1: Plant Species Clustering Around Sandalwood in Pachaimalai Hills

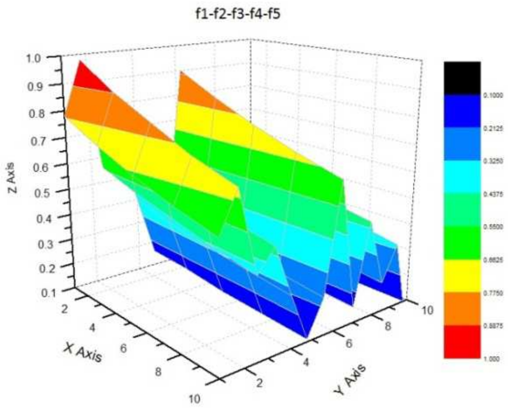

Figure 2: Three Dimensional Interpretation of the Comparison of the Plant Species

The fuzzy correlation coefficient is : $r_{f_1,f_2} = 0.98, \quad r_{f_1,f_3} = 0.99, \quad r_{f_2,f_3} = 0.95.$

Partial correlation is obtained as, $r_{1,(2\cdot3)} = 0.90, \quad r_{1,(3\cdot2)} = 0.95, \quad r_{2,(3\cdot1)} = -0.72.$

Semi-partial correlation is obtained as,

$$r_{1,(2\cdot3)} = 0.13, \quad r_{1,(3\cdot2)} = 0.19, \quad r_{2,(3\cdot1)} = -0.14, \quad r_{2,(1\cdot3)} = 0.28, \quad r_{3,(1\cdot2)} = 0.30, \quad r_{3,(2\cdot1)} = -0.10.$$

This multiple correlation coefficient is obtained as, $R_{1\cdot(2,3)} = 0.9982, \quad R_{2\cdot(1,3)} = 0.9902, \quad R_{3\cdot(1,2)} = 0.9957.$

REGRESSION ANALYSIS

Let us define $\omega = \begin{vmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{vmatrix}$ Where $r_{11} = r_{22} = r_{33} = 1.$ Expanding we get

$$\omega = 1 - r_{12}^2 - r_{23}^2 - r_{13}^2 + 2r_{12}r_{23}r_{13}.$$

Let ω_{ij} be the cofactor of r_{ij}, $\omega_{11} = \begin{vmatrix} r_{22} & r_{23} \\ r_{32} & r_{33} \end{vmatrix}, \quad \omega_{12} = -\begin{vmatrix} r_{21} & r_{23} \\ r_{31} & r_{33} \end{vmatrix}, \quad \omega_{13} = -\begin{vmatrix} r_{21} & r_{23} \\ r_{31} & r_{32} \end{vmatrix}.$

The equation of the plane of regression referred to as original origin is $\frac{X_1 - \overline{X_1}}{\sigma_1}\omega_{11} + \frac{X_2 - \overline{X_2}}{\sigma_2}\omega_{12} + \frac{X_3 - \overline{X_3}}{\sigma_3}\omega_{13} = 0$

Here, $\omega_{11} = \begin{vmatrix} 1 & 0.95 \\ 0.95 & 1 \end{vmatrix} = 0.0975,\ \ \omega_{12} = -\begin{vmatrix} 0.98 & 0.95 \\ 0.99 & 1 \end{vmatrix} = -0.04,\ \ \omega_{13} = -\begin{vmatrix} 1 & 0.98 \\ 0.99 & 0.95 \end{vmatrix} = 0.02$

Expanding we get, $\omega = 1 - (0.98)^2 - (0.95)^2 - (0.99)^2 + 2(0.98)(0.95)(0.99) = 0.00038$

Now $\sigma_1^2 = \frac{\sum X_1^2}{n} - \overline{X}_1^2 = 0.059,\ \ \sigma_2^2 = \frac{\sum X_2^2}{n} - \overline{X}_2^2 = 0.0499,\ \ \sigma_3^2 = \frac{\sum X_3^2}{n} - \overline{X}_3^2 = 0.0409.$

Since $\overline{X_1} = 0.54, \overline{X_2} = 0.49, \overline{X_3} = 0.45,$ now, $\frac{X_1 - 0.54}{0.24}(0.0975) + \frac{X_2 - 0.49}{0.22}(-0.04) + \frac{X_3 - 0.45}{0.2}(0.02) = 0.$

$X_1 = 0.4314 + 0.448 X_2 - 0.2462 X_3$ is the equation of regression plane of X_1 on X_2 and X_3.

RELIABILITY OF THE ESTIMATES

As for the two variable case, a measure of the reliability of the estimates of the dependent variable X_1 lies in the deviations of observed X_1 values from the estimated X_{1C} values based on the regression equation of X_1 on X_2 and X_3, as in above. These deviations $\left(X_1 - X_{1C}\right)$ are given in the table below:

Table 2: Estimated Values for Regression for Dependent X_1

X_1	X_2	X_3	X_{1C}	$\left(X_1 - X_{1C}\right)$	$\left(X_1 - X_{1C}\right)^2$
0.78	0.71	0.65	0.58945	0.19055	0.0363
0.98	0.89	0.82	0.628236	0.351764	0.1237
0.56	0.51	0.47	0.544166	0.015834	0.00025
0.66	0.6	0.55	0.56479	0.09521	0.00906
0.4	0.36	0.33	0.511434	-0.111434	0.01242
0.16	0.15	0.13	0.466594	-0.306594	0.094
0.34	0.31	0.28	0.501344	0.161344	0.026
0.9	0.82	0.75	0.61411	0.28589	0.0817
0.22	0.2	0.18	0.476684	-0.256684	0.0659
0.64	0.58	0.53	0.560754	0.079246	0.0063
0.3	0.27	0.25	0.049081	0.19081	0.0364
0.48	0.44	0.4	0.53004	-0.05004	0.0025

Here $\sum\left(X_1 - X_{1C}\right)^2 = 0.49453.$ A measure of the reliability of the estimates of the dependent variable X_2 lies in the deviations of observed X_2 values from the estimated X_{2C} values based on the regression equation of X_2 on X_1 and X_3, as in above. These deviations $\left(X_2 - X_{2C}\right)$ are given in the table below:

Table 3: Estimated Values for Regression for Dependent X_2

X_1	X_2	X_3	X_{2C}	$(X_2 - X_{2C})$	$(X_2 - X_{2C})^2$
0.78	0.71	0.65	1.133575	-0.423575	0.1794
0.98	0.89	0.82	1.67299	-0.78299	0.613
0.56	0.51	0.47	0.544065	-0.034065	0.00116
0.66	0.6	0.55	0.811025	-0.151025	0.0228
0.4	0.36	0.33	0.110335	0.289665	0.0839
0.16	0.15	0.13	-0.534765	-0.278235	0.0774
0.34	0.31	0.28	-0.05094	0.39094	0.15283
0.9	0.82	0.75	1.456125	-0.636125	0.4047
0.22	0.2	0.18	-0.37349	0.57349	0.3289
0.64	0.58	0.53	0.755435	-0.175435	0.0308
0.3	0.27	0.25	-0.156625	0.426625	0.182
0.48	0.44	0.4	0.3272	0.1128	0.0127

$\sum (X_2 - X_{2C})^2 = 2.08959.$

A measure of thereliability of the estimates of the dependent variable X_3 lies in the deviations of observed X_3 values from the estimated X_{3C} values based on the regression equation of X_3 on X_1 and X_2, as in above. These deviations $(X_3 - X_{3C})$ are given in the table below:

Table 4: Estimated values for regression for dependent X_3

X_1	X_2	X_3	X_{3C}	$(X_3 - X_{3C})$	$(X_3 - X_{3C})^2$
0.78	0.71	0.65	-0.1238	0.7738	0.5988
0.98	0.89	0.82	-0.6087	1.4287	2.0412
0.56	0.51	0.47	0.40595	0.06405	0.0041
0.66	0.6	0.55	0.1635	0.3865	0.1494
0.4	0.36	0.33	0.78295	-0.45295	0.2052
0.16	0.15	0.13	1.37575	-1.24575	1.5519
0.34	0.31	0.28	0.9357	-0.6557	0.4299
0.9	0.82	0.75	-0.4111	1.1611	1.3482
0.22	0.2	0.18	1.223	-1.043	1.0878
0.64	0.58	0.53	0.20835	0.32165	0.1035
0.3	0.27	0.25	1.0254	-0.7754	0.6012
0.48	0.44	0.4	0.60355	-0.20355	0.0414

$\sum (X_3 - X_{3C})^2 = 2.08959.$

The standard error of estimate, now denoted as is obtained from the sum of the squared deviations as

$$S_{1.23} = \sqrt{\frac{\sum (X_1 - X_{1C})^2}{N}},\ S_{2.13} = \sqrt{\frac{\sum (X_2 - X_{2C})^2}{N}},\ S_{3.12} = \sqrt{\frac{\sum (X_3 - X_{3C})^2}{N}}.$$

An alternate method of computation of $S_{1.23}$ in terms of simple correlation coefficients r_{12}, r_{13} and r_{23} is $S_{1.23} = S_1\sqrt{\frac{1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23}}{(1 - r_{23}^2)}}$; $S_{2.13} = S_2\sqrt{\frac{1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23}}{(1 - r_{13}^2)}}$; $S_{3.12} = S_1\sqrt{\frac{1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23}}{(1 - r_{12}^2)}}$. As before, $S_{1.23}$

represents the average difference between the observed values and the corresponding estimated X_{1C} values. The only difference being that X_{1C} values are now computed on the basis of the linear regression equations with X_1 as the dependent variable and X_2 and X_3 as two independent variables. The subscript 1 before the decimal point signifies the dependent variable X_1,and the subscript 2 and 3 after the decimal point signify the two independent variables X_2 and X_3.

When computing the standard error of an estimate $S_{1.23}$ based on the data given in the table comes to:

$$S_{1.23} = \sqrt{\frac{0.49453}{12}} = 0.203; \quad S_{1.23} = 0.27\sqrt{\frac{1-0.9604-0.9801-0.9025+2(0.98)(0.99)(0.95)}{1-0.9025}} = 0.0169$$

Similarlly, $$S_{2.13} = \sqrt{\frac{2.08959}{12}} = 0.4173; \quad S_{2.13} = 0.24\sqrt{\frac{1-0.9604-0.9801-0.9025+2(0.98)(0.99)(0.95)}{1-0.9801}} = 0.0332$$

and $$S_{3.12} = \sqrt{\frac{8.1625}{12}} = 0.8247; \quad S_{3.12} = 0.22\sqrt{\frac{1-0.9604-0.9801-0.9025+2(0.98)(0.99)(0.95)}{1-0.9604}} = 0.0216.$$

Least Square Method for Two Variable Cases

In the method of least square, the values of a and b are obtained by solving simultaneously the following pair of normal equations $\sum Y = aN + b\sum X; \quad \sum XY = a\sum X + b\sum X^2$ in which N is the total number of paired observations in the sample. The method of least squares seeks to minimize $W = \sum (Y - Y_C)^2; \quad W = \sum (Y - a - bX)^2$, with respect to a and b. Where $a = \frac{\sum Y \sum X^2 - \sum X \sum XY}{N\sum X^2 - (\sum X)^2}$ and $b = \frac{\sum XY - N\bar{X}\bar{Y}}{\sum X^2 - N\bar{X}^2}$.

A measure of the error of estimate is given by the standard error of estimate of Y on X, denoted as $S_{Y.X}$ and defined as $S_{Y.X} = \sqrt{\frac{(Y - Y_C)^2}{N}}$.

Table 5: Estimated Values for Least Square Regression

S. No	X_1	X_2	X_3	X_1^2	X_2^2	X_3^2	X_1X_2	X_2X_3	X_3X_1
1	0.78	0.71	0.65	0.6084	0.5041	0.4225	0.5538	0.4615	0.507
2	0.98	0.89	0.82	0.9604	0.7921	0.6724	0.8722	0.7298	0.8036
3	0.56	0.51	0.47	0.3136	0.2601	0.2209	0.2856	0.2397	0.2632
4	0.66	0.6	0.55	0.4356	0.36	0.3025	0.396	0.33	0.363
5	0.4	0.36	0.33	0.16	0.1296	0.1089	0.144	0.1188	0.132
6	0.16	0.15	0.13	0.0256	0.0225	0.0169	0.024	0.0195	0.0208
7	0.34	0.31	0.28	0.1156	0.0961	0.0784	0.1054	0.0868	0.0952
8	0.9	0.82	0.75	0.81	0.6724	0.5625	0.738	0.615	0.675
9	0.22	0.2	0.18	0.0484	0.04	0.0324	0.044	0.036	0.0396
10	0.64	0.58	0.53	0.4096	0.3364	0.2809	0.3712	0.3074	0.3392
11	0.3	0.27	0.25	0.09	0.0729	0.0625	0.081	0.0675	0.075
12	0.48	0.44	0.4	0.2304	0.1936	0.16	0.2112	0.176	0.192
$\sum$	**6.42**	**5.84**	**5.34**	**4.2076**	**3.4798**	**2.9208**	**3.8264**	**3.188**	**3.5056**

The regression equation of X_1 on X_2 is $X_{1C} = a + bX_2$.

Here, $a = \dfrac{\sum X_1 \sum X_2^{\,2} - \sum X_1 \sum X_1 X_2}{N\sum X_2^{\,2} - \left(\sum X_2\right)^2} = \dfrac{(6.42)(3.4798) - (6.42)(3.8264)}{12(3.4798) - 34.1056} = -0.2908.$

$b = \dfrac{\sum X_1 X_2 - N\bar{X}_1 \bar{X}_2}{\sum X_2^{\,2} - N\bar{X}_2^{\,2}} = \dfrac{3.8264 - (12)(0.49)(0.54)}{3.4798 - 12(0.49)^2} = 1.0879.$ Hence, $X_{1C} = -0.2908 + 1.0879 X_2$.

Table 6: Estimated Values for Least Square Regression for Dependent X_1

S. No	X_1	X_2	X_{1C}	$(X_1 - X_{1C})$	$(X_1 - X_{1C})^2$
1	0.78	0.71	0.481609	0.298391	0.089
2	0.98	0.89	0.677431	0.302569	0.0915
3	0.56	0.51	0.264029	0.295971	0.0876
4	0.66	0.6	0.36194	0.29806	0.0888
5	0.4	0.36	0.100844	0.299156	0.0895
6	0.16	0.15	-0.127615	0.287615	0.0827
7	0.34	0.31	0.046449	0.293551	0.0862
8	0.9	0.82	0.601278	0.298722	0.0892
9	0.22	0.2	-0.07322	0.29322	0.086
10	0.64	0.58	0.340182	0.299818	0.0899
11	0.3	0.27	0.002933	0.297067	0.0882
12	0.48	0.44	0.187876	0.292124	0.0853

$\sum\left((X_1 - X_{1C})^2\right) = 1.0539$, $S_{1.2} = \sqrt{\dfrac{\sum\left((X_1 - X_{1C})^2\right)}{N}} = 0.2964$

The regression equation of X_2 on X_3 is $X_{2C} = c + dX_2$

Here, $c = -0.2388$, $d = 1.1043$. Hence $X_{2C} = -0.2388 + 1.1043 X_3$.

Table 7: Estimated Values for Least Square Regression for Dependent X_2

S. No	X_2	X_3	X_{2C}	$(X_2 - X_{2C})$	$(X_2 - X_{2C})^2$
1	0.71	0.65	0.478995	0.231005	0.0534
2	0.89	0.82	0.666726	0.223274	0.0499
3	0.51	0.47	0.280221	0.229779	0.0528
4	0.6	0.55	0.368565	0.231435	0.0536
5	0.36	0.33	0.125619	0.234381	0.0549
6	0.15	0.13	-0.095241	0.245241	0.0601
7	0.31	0.28	0.070404	0.239596	0.0574
8	0.82	0.75	0.589425	0.230575	0.0532
9	0.2	0.18	-0.040026	0.240026	0.0576
10	0.58	0.53	0.346479	0.233521	0.0545
11	0.27	0.25	0.037275	0.232725	0.0542
12	0.44	0.4	0.20292	0.23708	0.0562

$\sum\left((X_2 - X_{2C})^2\right) = 0.6578$; $S_{2.3} = \sqrt{\dfrac{\sum\left((X_2 - X_{2C})^2\right)}{N}} = 0.2341.$

The regression equation of X_3 is $X_{3C} = e + fX_2$

Here, $e = 0.4042,\ \ f = 0.8323.$ Hence $X_{3C} = 0.4042 - 0.8323X_1.$

Table 8: Estimated Values for Least Square Regression for Dependent X_3

S. No	X_3	X_1	X_{3C}	$(X_3 - X_{3C})$	$(X_3 - X_{3C})^2$
1	0.65	0.78	1.053394	-0.403394	0.1627
2	0.82	0.98	1.219854	-0.399854	0.1599
3	0.47	0.56	0.870288	-0.4003518	0.1602
4	0.55	0.66	0.953518	-0.403518	0.1628
5	0.33	0.4	0.73712	-0.40712	0.1657
6	0.13	0.16	0.537368	-0.407368	0.1659
7	0.28	0.34	0.687182	-0.407182	0.1658
8	0.75	0.9	1.15327	-0.40327	0.1626
9	0.18	0.22	0.587306	-0.407306	0.1659
10	0.53	0.64	0.936872	-0.406872	0.1655
11	0.25	0.3	0.65389	-0.40389	0.1631
12	0.4	0.48	0.803704	-0.403704	0.163

$$\sum\left((X_3 - X_{3C})^2\right) = 1.9631;\ \ S_{3.1} = \sqrt{\frac{\sum\left((X_3 - X_{3C})^2\right)}{N}} = 0.4045.$$

Comparison of Least Square Method for Two Variable Cases with Three Variable Cases

Comparing $S_{1.2}$ with $S_{1.23}$, we notice that the magnitude of the variations in the two variable case is more than in three variable case. This means that the estimated X_{1C} values are closer to the observed values X_1 when the estimating regression equation has two independent variables than when it has one. This difference which represents increased accuracy is obviously due to the influence of the second independent variable X_3. In other words, it means that a part of the variations in X_1 which are unexplained by X_2 in the two variable cases are explained by the additional independent variable X_3 in the three variable cases. If the net influence of X_3 were zero, it will make no difference to the estimated X_{1C} values irrespective of X_3 being or not being in the estimating equation. But as long as there is a correlation between the dependent variable X_1 and any additional independent variable, the inclusion of the same in the regression estimating equation will reduce the error in the estimates of X_{1C} and improve their accuracy.

DISCUSSIONS

A random sample $T = \{t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8, t_9, t_{10}, t_{11}, t_{12}\}$ with 12 fuzzy records shown in Table 1 was initially taken at the beginning of this study, and $F = \{f_1, f_2, f_3, f_4, f_5\}$ was the set of observed fuzzy items which is the plant species mostly occurring in places where sandalwood is found. Fuzzy regression was computed, considering the three variables $\{f_1, f_2, f_3\}$. The whole study reveals a close relation and association between the variables in the set $\{f_1, f_2, f_3\}$.

It is identified that the plant species f_1, f_2, and f_3 are the most frequently occurring among the five total species under study (*Santalum album - Acacia pennata - Celtisphilippensis*). It is seen that the plant species *Acacia pennata* and *Celtisphilippensis*are very much closely associated than the other plant species identified in this study. A brief description of the botanical characteristics of the above-mentioned plant species is given as follows.

BOTANICAL NAME: *ACACIA PENNETTA* (*F_2*)

Family: *Mimosaceae*(Touch-me-not Family)

Climbing Acacia is a perennial climbing shrub or a small tree. The stem is thorny. Young branches are pubescent, green in color and turn brown with age. A large gland is present on the main spine of leaves above the middle of the petiole. Leaves are double-compound, pinnae 8-18 pairs, leaflets linear-oblong, smooth, base oblique or truncate, up to 50 pairs per pinna, hairy on the margins loosely set and overlapping. Flowers are in large panicles at the end of branches. They are spherical, pale yellow flower- heads. The pods are thin, flat and long with thick sutures.

Medicinal uses: In India, leaf juice mixed with milk is used for the treatment of indigestion in infants. It is also used for scalding of urine and for curing bleeding gums. Some people use boiled tender leaves for cholera treatment, digestive complaints, relief of a headache, body pain, snake bites, and even to cure fish poisoning. The root can be used for inducing flatulence and to cure stomach pain. The bark is used for the treatment of bronchitis, asthma and for stomach complaints.

BOTANICAL NAME: *CELTISPHILIPPENSIS* (*F_3*)

Family: *Cannabaceae*

Trees grow to about 12 m tall. The bark is usually grey, lenticellate and blaze brownish. The young branchlets are angular and pubescent. The leaves are simple, alternate and distichous; stipules lateral, pubescent, caducous and leaving scar; The petiole is 1 to 1.5 cm long, canaliculated above and pubescent when young. The lamina is 4.5 to 11.5 cm and elliptic to ellipticovate, the apex acute to acuminate, the base rounded or acute, sometimes asymmetric, the margin being entire. The flowers are inflorescence axillary cymes, polygamous and usually appear on new branchlets. The fruit is a drupe type up to 0.8 cm long and has one seed.

Economic Importance: The native Malayalis tribe people of Pachaimalai hills use the tree for firewood and for building construction purposes.

CONCLUSIONS

The proposed research work has dealt with various concepts related to fuzzy bio-statistical rules including regression analysis and reliability estimates to identify closely related plant species around sandalwood in the Pachaimalai hills, Tamilnadu, India. A roadmap to different correlation measures and regression analysis were clearly displayed in the numerical case study. The study revealed, after careful examination and computations, that two out of the four plant species are found to be very closely associated with sandalwood in Pachaimalai hills.

REFERENCES

1. *Agrawal, R, Imielinski, T& Swami, A. (1993).Mining Association Rules between Sets of Items in Large Databases, Proceedings of theACM SIGMOD International Conference on Management of Data, Washington D.C., pp.207-216.*
2. *Andreopoulou, Z., Manos, B., Polman, N., & Viaggi, D.(2011). Agricultural and Environmental Informatics, Governance & Management: Emerging Research Applications. Information Science Reference, ISBN: 1609606221, 9781609606220.*
3. *Chaudry, A. (2013). Forecasting rice production in West Bengal state in India: Statistical vs. Computational intelligence techniques.International Journal of Agricultural and Environmental Information Systems, 4(4), 68-91.*
4. *Han, J.,&Kamber, M. (2012).Data mining: Concepts and Techniques, Morgan Kaufmann Publishers. 3rd Edition, ISBN-10: 0123814790, ISBN-13: 978-0123814791.*
5. *Robinson, J.P., Chellathurai, S., &Prakashraj, G.D. (2012) A Strategic Study of Mining Fuzzy Association Rules Using Fuzzy Multiple Correlation Measures. Journal of Algorithms and Computational Technology, 6(3), 499-510.*
6. *Rojas-Mora, J., Josselin, D., Aryal, J., Mangiavillano, A., &Ellerkamp, P. (2013).The weighted fuzzy Barycenter: Definition and Application to forest fire control in the PACA region. International Journal of Agricultural and Environmental Information Systems, 4(4), 48-67.*
7. *Schleicher, J., Meyer, K.M., Wiegand, K., Schurr, F. M., & Ward, D. (2011).Disentangling facilitation and seed dispersal from environmental heterogeneity as mechanisms generating associations between savanna plants. Journal of Vegetation Science, 22(6), 1038-1048.*
8. *Selmaoui-Folcher, N., Flouvat, F., Gay, D., &Rouet, I. (2011).Spatial pattern mining for soil erosion characterization.International Journal of Agricultural and Environmental Information Systems, 2(2), 73-92.*
9. *Sonia, M.D., Robinson, J.P., and Rajasekaran, C.S., (2015). Mining Efficient Fuzzy Bio-statistical Rules for Association of Sandalwood in Pachaimalai Hills, International Journal of Agricultural and Environmental Information Systems, 6(2), 40-76.*
10. *Willner, W. (2006). The Association Concept Revisited. Phytocoenologia, 36(1), 67-76.*
11. *Yang, Y., Niu, Y., Cavieres, L, A., & Sun, H. (2010). Positive associations between the cushion plant* Arenariapolytrichoides*(Caryophyllaceae) and other alpine plant species increase with altitude in the Sino-Himalayas. Journal of Vegetation Science, 21(6), 1048-1057.*
12. *Zadeh, L.A. (1978). Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems, 1,3-28.*
13. *http://www.britannica.com/EBchecked/topic/437501/Pachaimalai-Hills*
14. *http://wikimapia.org/9501280/PACHAMALAI-HILLS*